Added Value Revisited: On the Sound/Image Interaction in Film

Flückiger, Barbara

Added Value Revisited
On Sound/Image Interaction in Film

Barbara Flückiger

—

Barbara Flückiger builds on concepts established by Michel Chion, particularly those of "synchresis" and "added value," in order to re-evaluate the interaction between sound and image in film. The historical development of sound in relation to the filmic image, the constructedness of soundtracks, and the subjects of non-simultaneity and immediacy in relation to the production of movie soundtracks constitute the main lines of Flückiger's essay, along with an investigation into the historical development of specific vocabularies in film sound.

Barbara Flückiger

Added Value Revisited
On Sound/Image Interaction in Film

—

On the subject of the sound/image relationship in film, Michel Chion's *L'Audio-Vision*[1] remains the most valuable source of fundamental thought to date. In particular, the concepts of *synchresis* and *added value* that he discusses there are useful when discussing the role human perception has in the combination of visual and acoustic information presented by the cinematic work of art. However, analysis of sound/image interaction could be developed even further. How exactly do sound objects transform visual perception? And what is the difference between this interaction in everyday life and in film? What conditions are necessary to allow for a coherent connection between the visible object and the sounds related to it on a film soundtrack?

As Michel Chion has pointed out, there is a widespread belief even among film scholars that sound is somehow naturally there, simply emanating from the depicted space, its objects and the characters populating it. In fact, however, there is a long history of sound substitution that dates back to ancient Greek theater. As far back as the first century BC, a Greek technician – Heron of Alexandria – had already devised hydraulic machines for the production of sound effects such as wind, thunder, or birds chirping. This tradition was pushed further in the theater technology of the Baroque era. Although sound effects may have been largely absent in the silent cinema, there was already a broad knowledge of artificial sound production as demonstrated in a French publication from 1914, S. de Serk's *Les Bruits de coulisses au cinéma*. De Serk describes many techniques, including the use of coconut shells to substitute for horses' hooves – a technique used by Foley artists for decades. From the early days of sound cinema, it was clear to sound engineers that merely recording what happens on the set, deriving from the events taking place in front of the camera, would not guarantee

convincing results. Out of that knowledge, a sound production process emerged which in the Hollywood system led to the practice of mostly recording dialogue, adding all the other sound elements and constructing the sound/image relationship afterwards through editing and re-recording. Moreover, every sound technician knows that most sound objects trigger the correct mental image only when they are carefully constructed. Probably the best example of this phenomenon is the sound of gunshots, which have established their own distinctive shape in the history of sound cinema. This acoustic shape differs considerably not only from actual sound recordings but also from the sound of gunshots as they are heard by the human ear in the real world. Therefore the reference for the sound perception that informs the expectations of viewers/listeners is not grounded in their everyday life experience but in their experience of deciphering audiovisual codes.

Besides the culturally determined roots of sound production's traditions, there are also technical factors behind this development. For decades, sound recording and reproduction equipment were marked by the poor definition of the monaural optical track. These technical shortcomings affected the representation of sound effects much more than the recording of music and voice, because of the non-harmonic structure of the former. In perceiving musical sound and spoken vowels, which have a harmonic structure, the human ear is capable of filling in those frequencies that are lost in the process – but this cannot occur when complex acoustic events are perceived. This psychoacoustic consideration led to the dominance of certain sound effects with a well-defined acoustic shape such as car horns, church bells and animal voices. Although one might expect that the aural identification of a sound object would be of significance only when a sound is presented as acousmatic, i.e. without the synchronous presentation of a visual source, in the Hollywood classical era it was an unwritten rule never to use any sound that the audience cannot grasp immediately.

As a result, cultural aspects and technical limitations generated a rather clear-cut sound vocabulary. It was not until the mid-1970s that both cultural changes in the wake of the 1960s and a universally available technical improvement – the Dolby system – paved the way for the introduction of a broader diversity of sound effects and more complex soundtracks. However, it should be noted that the magnetic sound systems that emerged with the wide-screen formats in the 1950s offered an even better tonal and spatial definition than the optical Dolby system introduced in the 1970s. Therefore the lack of significant change in the 1950s can be attributed to the cultural predominance of the Hollywood production mode and additional cultural factors.[2]

Beyond these fundamental observations regarding the development of specific vocabularies in the history of film sound, a semio-pragmatic aspect should be taken into account when investigating the reception of fiction films, namely the intention that is at work in these artifacts. The viewers/listeners in the cinema are well aware that every detail of the movie is carefully selected in order to contribute to the narrative or, in other words, that every detail has to make sense and is not left to chance as in the perception of our everyday environment. Never, for example, would a gunshot be heard when the loving couple is lying in bed on a Sunday morning, unless such a sound effect were meant to announce an upcoming danger and therefore a plot point in the story construction. Never would a door open in the background without announcing the arrival of another character, be it a friend or an enemy. Thus the cultural frame in which the sound/image relationship occurs strongly shapes its perception.

Multiple Constraint Satisfaction Theory

Once these fundamental considerations are set we can take a closer look at the interaction of visual and acoustic clues in the cinematic framework. The term *added value* designates an energetic flux between two concepts leading to a third one, which is contained in neither. Therefore the effect is a process of modification, similar to the semantic exchange in metaphors. This interaction leads to the emphasis of certain aspects of the objects – for example their materiality – while at the same time suppressing others. In his talk at the "Utopia of Sound" conference in Vienna, Michel Chion showed an excerpt from Andrei Tarkovsky's *Solaris* (USSR 1972) where the movements of a woman's body are accompanied by the sound of glass, thereby evoking the impression that this body has changed its materiality from human flesh to glass without any change in the visual representation. This is indeed a striking case of transformation, which may be regarded as a limit for what is possible without causing the sound/image relationship to break down. One should note, however, that there is an explanation for this effect in the narrative, which leads the viewer to accept this change as plausible. Similar cases occur in the tradition of the animated film, where both the stylized image and the sound are marked by a great degree of divergence from the familiar experiences of everyday life. This remoteness from a reality effect offers an unusual artistic freedom that defies any physical law known through our observation of and interaction with the world.

On the other hand, there is redundancy where no transformation and therefore no added value could be found. I would challenge Chion's notion that redundancy is impossible. To this purpose it is useful to stress the fact that sounds carry several layers of information, which can be divided into sensory and semantic forms. Spoken language consists of both semantic information – the content of the speech – and sensory elements, which are called the prosodic elements of language. These acoustic elements convey information about the emotional state of the speaker as well as about his or her social and ethnic background. Furthermore, the tone of the voice stems in part from the material conditions of the body, and is quite a reliable indicator of the speaker's sex and – to a lesser degree – of his or her age and physical condition. The same holds true for any sound object. A door closing can mean that someone is coming or that someone is closed in. As I have shown in a case study of *The Silence of the Lambs* (Jonathan Demme, USA 1991), the sound of doors even carries with it a symbolic layer of meaning that stems from the door as an object and its use in crossing borders between spaces or its use in locking away unwanted members of society.[3] When listening to the sensory part of door sounds it becomes evident that they bear information about the material and the shape, about its weight and condition. Squeaking doors – as a stereotypical element of some genres – turn a sensory aspect into higher-level information whose meaning is based on an intertextual foundation closely linked to a definite set of associations that trigger instant emotional responses.

In my study on sound design I formulated the following hypothesis:[4] The more elaborate a sound object is on the semantic and sensory level, the greater the added value resulting from its interaction with the depiction of the hypothetical sound source. Therefore a very standardized sound object with a poor definition has the least impact on visual representation. Its meaning might not surpass a simple deixis towards the object, which could be expressed as "there is a ...," while the sensory layers lack any specific traits. That is when redundancy occurs. Take the notorious sound of coconut shells produced by the Foley artist as a substitute for the tapping of horses' hooves. From the thousands of possible sounds that could stem from the interaction of the hoof with any type of material it equals a featureless average. In a classical Hollywood film the depicted horse would be typified in a similar fashion. In such a case, the visual and the acoustic representation overlap completely while they – of course – address two different senses. However, since there is no effect of transformation on the visual perception the added value tends to zero.

On the other hand, when we take a very distinct sound object like the sound of the laser sword in the *Star Wars* saga, a complex interaction is taking place. Before we see the object for the first time, the uncle Ben Kenobi tells Luke Skywalker – and us – that he is going to present the "father's light saber." What we perceive visually, though, looks like a short metal stick that glows like a neon tube when turned on – obviously lacking the attribute "dangerous" that would be expected from a weapon. Sound designer Ben Burtt had the task of creating a sound object that, by its very sensory qualities and semantic associations, would enhance the visual appearance in the intended way. Burtt, being a trained scientist, started with the hum of a broken television set, which he recorded using an interference microphone. By moving the microphone in sync with the movement on the screen he produced a changing humming sound that was accompanied by phasing. Phasing occurs when an interference microphone moves in a diffuse sound field of broadband noise, and is normally an indicator of very fast movements such as those produced by jet planes.

Thus the construction of the object is achieved by integrating a set of disparate characteristics, and therefore far surpasses the simple indicative function of the horse example. Specific characteristics were used to establish functions including consequences, processes, a hypothetical material composition and sensory qualities. The sound object itself is composed of various component concepts, which are associated with various processes: the tone modulation of the humming sound as the object moves in space, the hissing sound as it is switched on, the crackling of contact. Furthermore, the tonal qualities address familiar sense experiences from everyday life: the association with speed and power by means of phasing as well as negative experiences of electricity and heat by means of a crackling sound.

When we look closer at the functioning of the so called cross- or inter-modal association that is necessary for the connection of auditory and visual stimuli in human perception, we can gain an even deeper insight into the structure of added value. In the process of deciphering the cinematic *polyphonie informationelle* – as Roland Barthes has called it – the viewer/listener brings his or her capacity into play to connect the different aspects of the world represented in the film. As the psychophysicist Hermann Helmholtz pointed out in the second half of the nineteenth century, the sense modalities are focused at specific bands of energy only. There is no continuum between what he called *Qualitätenkreise*:[5] the sensorial qualities of the seen, the heard, the touched, etc. Even before Helmholtz, the biologist Johannes Müller[6] had developed an even more radical concept which he called *Prinzip der undifferenzierten Kodierung* (*principle*

of undifferentiated coding or *law of specific sense energies*). This principle states that all sense receptors code the intensity of the stimuli only, without regard to the physical or chemical nature of their source. That means they code only "so-and-so-much" at this point in my body and not "what."[7] As a consequence of this observation we can understand that our perception is fundamentally fragmented. By dividing representation into visual and acoustic aspects, the cinematic apparatus thus only mimics this fragmentation.

This is based on the simultaneous presentation of visual and auditory stimuli that leads to the assumption of a causal link between them. Michel Chion coined the term *synchresis* for the irresistible and spontaneous bonding that glues the two forms together. Furthermore this phenomenon is the very foundation of any Gestalt perception that is grounded in the synchronous firing of neural cells. The neurologist Wolf Singer[8] has stated that our knowledge of consistent relations between phenomena in the world is based on the repeated simultaneous firing of neural groups that – as a result – form new connections. The more these connections are activated the stronger they become. Today it is assumed that these solidified neural networks constitute the mental representations in long-term memory and are the prerequisite of any identification of objects. But there are also inhibitory processes that govern the perception of stimuli with contradictory qualities such as round and sharp or high- and low-pitched, which never occur at the same time. This observation has led to the "theory of multiple constraint satisfaction,"[9] which declares that there are limits to what kinds of attributes can be summed up to the perception of an integral whole by the human brain.

When we now reconsider the concept of value added, we can conclude from the neuropsychological facts mentioned above that this value equals the effort one has to invest in perceiving the connection between the different concepts. Hardwired relationships such as the horse/coconut shell association described earlier are perceived instantly and automatically. They do not require any attention. More complex relationships with ambiguous attributes on both sides of the audiovisual representation call for much more investment in the course of their perception. Still, there is a limit to what is possible when these relationships contradict the knowledge acquired in everyday experience. This is one – perhaps the most important – reason why the contrapunctal use of sound so dear to early film theoreticians like René Clair or Sergei Eisenstein has not found widespread use. In any case, the most critical basis for connecting sound and image are the material foundations: the tone evoked by a material – such as wood, metal or water – and the mass as well as the volume of the sound-generating source. If

these foundations are disregarded an alienating effect is generated, as shown in the example from *Solaris*.

As demonstrated in the example of the laser sword, emotions arise from associations established by tradition and/or the personal, subjective memories they trigger. It is clear of course that American mainstream films have to be much bolder in their use of expressive means than European films, because they aim at a global market and therefore are to be understood by a very heterogeneous audience with different cultural backgrounds. The most sophisticated films and the best sound designers pay attention to this requirement while at the same time offering additional layers that lead to more subtle responses in certain parts of the audience – as can be shown in the close analysis of these films.

However, there are certain exceptions to the constraint satisfaction theory. One of them is a genre like the animated film, which – as mentioned before – has established a stylized repertoire of its own. In a similar fashion the self-reflexive comedy builds on the vaudeville tradition of working with exaggerations.

One can conclude from these obvious examples that a semio-pragmatic framework guides the expectations and therefore the reception of the audience and allows for greater artistic freedom and greater deviation from sound/image relationships as experienced in the real world. Thus the fiction functions like a filter that allows the audiovisual artifact to transform natural perception. This effect is in perfect accordance with the process of defamiliarization (*ostranenie*) called for by the Russian Formalist Viktor Shklovsky. In his view the difference quality of the artwork should interrupt automated processes of perception, thereby sharpening the senses. However, this effect fades when traditions lead to new standardizations, thereby stereotyping the means of artistic expression.

In addition to this fundamental function of the work of art, some specific narrative strategies allow for stronger transformations. When the diegesis, or part of it, is presented from the perspective of a character, this subjective perception is generally presented as a dissociation of sound and image, visual and auditory cues evidently deviating from one another. Spectators are forced to draw their conclusions from the unusual mode of presentation and to attribute it to a shift in what Gérard Genette[10] called *focalization*, the shift from a non-focalized – i.e. a "neutral," auctorial perspective – to an internally focalized – i.e. a subjective – perspective. A sudden increase in reverberation that is not explained by a spatial change, slow motion accompanied by distorted sound effects, clearly augmented sound objects, and internal sounds such

as breathing and heartbeats are some of the most common strategies for the representation of subjective transformations. All of these strategies are more or less rooted in the psychological phenomena of sensorial dissociation occurring in dreams, situations of stress or even psychotic episodes. Thus the shift in the sound/image relationship oscillates between the mimicking of natural perception and the coded narration of such a displacement. While in previous periods of film history these shifts were clearly marked by an introduction such as a tracking shot or a zoom into the face or the eye of the focalizer – i.e. the character whose perception we are going to experience – today many films draw their fascination from an ambiguous play with several layers of reality. More and more, viewers/listeners are confronted with complex arrangements of the audiovisual medium. As a result, even in their consciousness the sound objects lose their once-subordinate status, becoming an independent means of expression instead of simply emanating from the images on the screen.

—

1 Michel Chion, *L'Audio-Vision*, trans. Claudia Gorbman (New York: Columbia University Press, 1994).

2 This notion is based on my careful analysis of 96 films produced between 1926 and 1995 in the course of my research project "Sound Design in American Mainstream Film from 1975 to 1995." The results were published as *Sound Design. Die virtuelle Klangwelt des Films* (Marburg: Schüren) in 2001, 3rd edition 2006.

3 Barbara Flückiger, *Sound Design* 173 f.

4 Flückiger 143 ff.

5 Hermann Ludwig Ferdinand von Helmholtz, *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik* (Braunschweig: Vieweg, 1862) 13. [in English: *On the Sensations of Tone as a Physiological Basis for the Theory of Music*, trans. Alexander J. Ellis (London: Longmans, Green, 1875)].

6 Johannes Müller, *Handbuch der Physiologie des Menschen für Vorlesungen* (Koblenz: J. Hölscher, 1837-1840).

7 Heinz von Foerster, "Wahrnehmen" (1988), *Philosophien der neuen Technologie*, ed. Ars Electronica (Berlin: Merve Verlag, 1989) 35.

8 Wolf Singer, "Der Beobachter im Gehirn," *Der Mensch und sein Gehirn. Die Folgen der Evolution*, ed. Heinrich Meier and Detlev Ploog (Munich, Zürich: Piper, 1997) 53.

9 Thomas Goschke and Dirk Koppelberg, "Konnektionistische Repräsentation, semantische Kompositionalität und die Kontextabhängigkeit von Konzepten," *Interdisziplinäre Perspektiven der Kognitionsforschung*, ed. Helmut Hildebrandt and Eckart Scheerer (Frankfurt: Lang, 1993).

10 Gérard Genette, *Figures III* (Paris: Editions du Seuil, 1972), and *Nouveau discours du récit* (Paris: Editions du Seuil, 1983) [in English: *Narrative Discourse Revisited*, trans. Jane E. Lewin (Ithaca NY: Cornell University Press, 1988)].